M.(ϕ)BssHII, a novel cytosine-C5-DNA-methyltransferase with target-recognizing domains at separated locations of the enzyme

Svenja Sethmann¹, Piotr Ceglowski²,
Jürgen Willert, Roksana Iwanicka-Nowicka²,
Thomas A. Trautner³ and Jörn Walter

Max-Planck-Institut für Molekulare Genetik, Ihnestr. 73, 14195
Berlin, Germany

¹Present address: Deutsches Rheumaforschungszentrum, Hannoversche
Straße 27, D-10115 Berlin, Germany

²Permanent address: Institute of Biochemistry and Biophysics, Polish
Academy of Sciences, 5a Pawinskiego St, 02-106 Warszawa, Poland

³Corresponding author
e-mail: Trautner@mpimg-berlin-dahlem.mpg.de

In all cytosine-C5-DNA-methyltransferases (MTases) from prokaryotes and eukaryotes, remarkably conserved amino acid sequence elements responsible for general enzymatic functions are arranged in the same canonical order. In addition, one variable region, which includes the target-recognizing domain(s) (TRDs) characteristic for each enzyme, has been localized in one region between the same blocks of these conserved elements. This conservation in the order of conserved and variable sequences suggests stringent structural constraints in the primary structure to obtain the correct folding of the enzymes. Here we report the characterization of a new type of a multispecific MTase, M.(ϕ)BssHII, which is expressed as two isoforms. Isoform I is an entirely novel type of MTase which has, in addition to the TRDs at the conventional location, one TRD located at a non-canonical position at its N-terminus. Isoform II is represented by the same MTase, but without the N-terminal TRD. The N-terminal TRD provides *Hae*II methylation specificity to isoform I. The TRD is fully functional when engineered into either the conventional variable region of M.(ϕ)BssHII or the related monospecific M. ϕ 3TII MTase. The implications of this structural plasticity with respect to the evolution of MTases are discussed.

Keywords: enzyme domains/enzyme isoforms/molecular evolution/protein–DNA interaction

Introduction

Methylation of DNA-cytosines at their C5 position plays a key role in the maintenance of structural integrity and the realization of the genetic information in prokaryotes and eukaryotes (Noyer-Weidner and Trautner, 1993; Yoder and Bestor, 1996). Methylation is mediated by C5-DNA-methyltransferases (MTases). Depending on the number of targets recognized by an MTase, MTases may be mono- or multispecific. The basic mechanism of methylation is the same in all organisms studied (Bestor and Verdine, 1994). It involves the transfer of a methyl group from

S-adenosylmethionine (SAM) to the C5 position of the recipient DNA-cytosine with the release of one proton and the generation of methylated DNA and S-adenosylhomocysteine. The universality of this mechanism is paralleled by the extreme phylogenetic conservation of MTases (Lauster *et al.*, 1989; Kumar *et al.*, 1994). In all enzymes irrespective of their phylogenetic origin, 10 conserved amino acid motifs are arranged in a canonical order, separated by sequences variable in size and amino acid composition in different enzymes.

Crystallographic studies on two bacterial MTases showed that general enzymatic functions such as cofactor (SAM) binding and catalysis are provided by a large ‘catalytic’ domain formed by the N-terminal part of the enzyme which includes the conserved motifs I–VIII. The second, smaller ‘recognition’ domain of the MTases is composed of the central variable region (V) containing the target-recognizing domain(s) (TRDs) and part of the C-terminal motifs IX and X (Klimasauskas *et al.*, 1994; Reinisch *et al.*, 1995). Motif X forms one of the bridges between the two domains. In prokaryotic MTases, the two-domain structure represents the MTase holoenzyme. In most eukaryotic MTases, however, this structure makes up only about one-third of the holoenzyme. Such MTases have an additional extensive N-terminal structure, which confers regulatory functions controlling general parameters of methylation such as developmental regulation, interaction with other proteins and proper localization in the cell (Leonhardt *et al.*, 1992; Bestor and Verdine, 1994).

The invariant succession of conserved and variable regions in all known primary structures of MTases suggests stringent structural constraints for the appropriate interactions of the two domains to catalyze sequence-specific methylation. In particular, the integrity of the small domain, providing a correct interplay between the variable TRD-containing region and the conserved motifs IX and X, appears to be crucial for the enzymatic activity. This assumption followed experiments in which chimeric monospecific MTases were analyzed (Mi and Roberts, 1992). Activity of such chimeric MTases was only observed when the entire small and large domains were exchanged, whereas an exchange of only the variable region resulted in inactive enzymes.

Here we present the characterization of a MTase, M.(ϕ)BssHII, with a new primary structural architecture, showing that there are alternatives to integrating a TRD in the enzyme. This multispecific MTase exists as two isoforms: one is a conventionally structured MTase with TRDs only in the variable region connected to the small domain. The second form of this enzyme has, in addition, an extended N-terminus containing an extra TRD at this non-canonical location outside of the conventional variable region. We discuss our results in terms of the structural requirements for methylation and the evolution of other MTases, particularly eukaryotic MTases.

Fig. 1. Structure of the *M.(\phi)BssHII* gene, its sequence of the 5' region and amino acid sequence alignment of TRDs. (A) The upper part shows the overall structure of the *M.(\phi)BssHII* gene with the ORFs depicted as filled bars and the untranslated regions as thin bars. The TRDs are highlighted by black shading and numbers. TRDs 1 and 4 are shown in parentheses as no methylation specificity has yet been assigned to them. The extents of the two isoform ORFs starting at ATC and ATG are indicated above as horizontal lines. The lower part gives the nucleotide and deduced amino acid sequence of the 5' region of the *M.(\phi)BssHII* gene. Nucleotide and amino acid sequence numbers are shown on the left and right, respectively. Nucleotides in lower case letters refer to sequence corrections of the previously published sequence detected in the course of this work (DDBJ/EMBL/GenBank accession No. X90873, corrections were submitted to GenBank). The start codons for isoforms I and II, ATC₁₄₉₋₁₅₁ and ATG₃₄₄₋₃₄₆ are boxed. The ORFs of isoforms I and II are marked by light and heavy underlining. The transcriptional start sites of the *M.(\phi)BssHII* gene at A₇₁ and A₇₅ mapped by primer extension (data not shown) are marked by vertical arrows. The sequence (A₁₅₄-G₂₆₂) deleted to construct *M.(\phi)BssHII* Δ_0 (see Figure 3) is indicated by broken arrows. The position of the point mutation C₂₅₉T (Mut₃₈) is labeled by a *. The amino acid sequence blocks shown in black (NYD) and (HP) confine the fragment used for amino acid sequence alignment. (B) Amino acid sequence alignment of TRD0 and TRDs 2-5 from the variable region. Identities to TRD0 are shaded, the conserved flanking motifs N(H,Y) (D,N,E), (H,Y)P and PXXT(I,M) are boxed. These motifs confine TRDs 2-5 as described in Schumann *et al.* (1996) and as determined by mutational analysis (D.Koch, unpublished).

Results

From our previous studies, the primary structure of the multispecific MTase *M.(\phi)BssHII* followed the canonical build up of MTases with sequentially arranged conserved and variable regions (Figure 1A). The enzyme methylates five DNA target sequences (Schumann *et al.*, 1996). Within the central variable region of this MTase, TRDs responsible for recognition of the sequences AC*GCGT and CC*GCGG (*Mlu*I and *Sac*II, TRD2), PuC*CGGPy (*Cfr*10I, TRD3) and GCGC*GC (*Bss*HII, TRD5) have been identified by genetic analysis. Surprisingly, the TRD recognizing the PuGC*GCPy (*Hae*II) target could not be localized in this variable region. A deletion of the entire variable region containing the identified TRD elements of the *M.(\phi)BssHII* enzyme [*M.(\phi)BssHII* Δ_{1-5}], did not affect its potential to methylate *Hae*II sequences (Schumann *et al.*, 1996; see also Figure 3). This finding suggested that the *Hae*II-specific TRD, which we term TRD0, had to be located at an

unusual position outside of the canonical variable region. By a series of biochemical and genetic experiments, we show that the TRD0 is indeed located at an unusual position N-terminal to the ensemble of conserved elements.

M.(\phi)BssHII exists as two isoforms with different methylation capacities

Our previous analyses suggested that the *M.(\phi)BssHII* protein represents a single polypeptide of ~59 kDa. Western blot analysis of *M.(\phi)BssHII* proteins separated on low percentage SDS-PAGE, however, revealed the presence of two isoforms of the enzyme. In addition to a 59 kDa polypeptide (isoform II), the MTase is translated into a second, larger protein of apparently 66 kDa (isoform I) (Figure 2A and B). In both *Escherichia coli* and *Bacillus stearothermophilus* cells expressing the *M.(\phi)BssHII* gene, the amount of isoform I is greatly reduced in comparison with the predominant isoform II. On determining the N-terminal sequence of both purified isoforms (see

Fig. 2. Identification and methylation potential of *M.(phi)BssHII* isoforms. (A) Western blot of protein extracts from *E.coli* cells expressing *M.(phi)BssHII*. Left lane: the two isoforms cross-reacting with a polyclonal *M.(phi)BssHII* antibody. Right lane: molecular weight standard. (B) PAGE of purified *M.(phi)BssHII* WT (left lane) and purified His-tagged isoforms I and II (right lanes). The same molecular weight standards were used in (A) and (B). (C) Agarose gel analysis of plasmid pMS119EH DNA *in vitro* methylated with purified His-tagged isoforms I or II followed by digestion with the restriction enzymes indicated above ('-', undigested; B, *R.BssHII*; H, *R.HaeII*; Cf, *R.Cfr10I*; S, molecular weight standard VII of Boehringer Mannheim).

Figure 2B), we realized that they are translational products of the same open reading frame (ORF). Isoform II is initiated at ATG_{344–346}, and the newly identified isoform I at the unusual codon ATC_{149–151}, some 79 codons 5' of the translational start of isoform II (Figure 1A). Both isoforms are most probably alternative translational products of the same mRNA, as transcription of the *M.(phi)BssHII* gene initiates 5' to these translational starts at A₇₅ and A₇₉ as mapped by primer extension analysis (see Figure 1A; data not shown). The relative amounts of enzyme produced must be regulated at a post-transcriptional level. This control, however, does not depend primarily on the translational signal sequences preceding

the genes, since several mutagenesis experiments designed to affect the expression of one or the other isoform did not lead to readily interpretable changes (data not shown).

The identification of two translational products of the *M.(phi)BssHII* ORF raised the question of their methylation potential, in particular whether the two forms of the enzyme differed in their capacities to methylate *HaeII* sequences. To monitor this, the proteins were expressed as N-terminal fusions at ATC_{149–151} or ATG_{344–346} to histidine tags and purified to homogeneity (Figure 2B). The addition of the histidine tags did not interfere *in vivo* with the activity of either isoform (Figure 2C). Methylation at *HaeII* sequences was only performed by isoform I. Such evidence for the differential activity of the two isoforms came also from *in vitro* analysis of the affinity-purified proteins (Figure 2C). Together, these results indicated that the capacity of *M.(phi)BssHII* to methylate *HaeII* sequences is correlated with the presence of the extended N-terminal part of isoform I, suggesting that a *HaeII*-specific TRD0 is located in this part of this protein.

Support for this notion came from experiments in which sequences N-terminal to the translational start of isoform I of *M.(phi)BssHII* and the highly related MTase *M.(phi)3TII* (Figure 3) were exchanged reciprocally. *M.(phi)3TII* is a monospecific MTase which has no TRD0 equivalent and specifies only TC*GA methylation by a TRD T located within the canonical central variable region of the enzyme. Addition of the N-terminal sequences of *M.(phi)BssHII* to *M.(phi)3TII* generated chimera I which exhibits both *HaeII* and *TaqI* methylation capacities, whereas the reciprocal chimera II exhibits all genuine *M.(phi)BssHII* methylation specificities except that for *HaeII* sequences (Figure 3).

Moreover, by inspecting the derived amino acids of the N-terminal sequence of isoform I, we identified a PXXTI motif characteristic for TRDs (Lauster *et al.*, 1989; Kumar *et al.*, 1994; Schumann *et al.*, 1996). An amino acid alignment at this motif revealed a striking similarity to various TRDs of the variable region of *M.(phi)BssHII* (Figure 1B). Based on this alignment, TRD0 has 52 amino acids assuming that the motifs NYD and HP constitute its N- and C-terminal confines. Similar motifs have been shown experimentally to demarcate the extent of other TRDs of the variable region of this enzyme (Schumann *et al.*, 1996; D.Koch, unpublished).

Mutations within TRD0 destroy the *HaeII* methylation capacity

To confirm the location of TRD0, we generated mutations in the proposed TRD0-encoding region of the *M.(phi)BssHII* gene. In two deletion constructs, in which a substantial part of the TRD0-coding sequences but not the start codon ATC_{149–151} were eliminated (*M.(phi)BssHII* Δ_{1-5} ; Figure 3), the *HaeII*-specific methylation capacity was lost. The activities specified by the other TRDs remained unaffected. Furthermore, a C₂₅₉T missense mutation (Figure 1A), leading to a P₃₈S amino acid change within TRD0, had the same phenotype (not shown, see also chimeras IV_{mut38} and V_{mut38} in Figures 3 and 4). Together with the *M.(phi)BssHII* Δ_{1-5} deletion, these mutations provide direct genetic evidence for the location of TRD0 at the proposed position.

Fig. 3. Structure and functional properties of M.(ϕ)BssHII and M. ϕ 3TII, and their chimeric and deletion derivatives. The ORFs of the M.(ϕ)BssHII (filled) and M. ϕ 3TII (white) genes are given above. Boxes labeled I–VIII and IX–X define the relative location of the two blocks of conserved elements flanking the variable TRD region. The TRD equivalents are shown as boxes 0, 1–5 and T. All genes are aligned at the location of a unique BglII (Bg) site used as a chimeric fusion point. In all chimeric MTases, a mutated, inactive TRD5 (indicated by an asterisk) was used, since this TRD exhibited some relaxed methylation capacity in a non-wild-type context which interfered with a clear assignment of the methylation specificities of the other TRDs. The methylation capacity of the parental and derived MTases were tested by the sensitivity of MTase-expressing plasmid DNA against the cognate restriction endonucleases R.HaeII (Hae), R.MluI (Mlu) and R.Cfr10I (Cfr). To test 5' TCGA3' (Taq)-specific methylation, plasmids were digested with R.ClaII (5' ATCGAT3') or R.SalI (5' GTCGAC3'). The results of these analyses are schematically depicted in the right panel of the figure ('+', protection of DNA against the cognate restriction endonuclease; '-', no protection; '+/ -', partial protection).

TRD0 constitutes a transferable TRD module

The activity of chimera I, in which the N-terminus of the M.(ϕ)BssHII isoform I was fused to the M. ϕ 3TII ORF, revealed that TRD0 is also functional when combined with the enzyme core of a related MTase. In order to test whether TRD0 indeed functions as an autonomous recognition element, we constructed another series of chimeras in which TRD0 was placed in various positions and sequence contexts of both M.(ϕ)BssHII and M. ϕ 3TII. In one of these chimeras (chimera III), the entire variable region of M.(ϕ)BssHII (i.e. TRDs 1–5) was substituted by

a TRD0 segment confined by the NYD and HP motifs (see Figure 1B). Chimera III exhibited exclusively HaeII-specific methylation activity (Figure 3). Two conclusions can be drawn from this result: (i) TRD0 is fully functional in the proposed confines and (ii) this TRD confers sequence-specific recognition not only in its original structural context at the N-terminus of the protein but also within the canonical position of TRDs. Moreover, the phenotypes of chimeras IV and V (Figures 3 and 4) demonstrate that TRD0 cannot only be functionally integrated into the variable region by substitution for existing

Fig. 4. Methylation analysis of plasmids expressing parental and chimeric MTase genes. The agarose gel shows restriction endonuclease-digested plasmids encoding the M.ϕ3TII gene and the chimeras IV and V harboring TRD0 or its mutated form (Chimera IV_{Mut38} and Chimera V_{Mut38}), respectively (see Figure 3). Lanes are labeled according to the restriction endonuclease used: '–', undigested; H, *R.HaeII*; Cf, *R.Cfr10I*; M, *R.MluI*; C, *R.ClaI*; E, *R.EcoRI*; E/C, double digestion with *R.EcoRI* and *R.ClaI*; S, molecular weight standard (Ladder Mix, MBI Fermentas). For a proper examination of digestion with *R.ClaI*, the plasmids, which contain only one restriction site for this enzyme, were double digested with the methylation-insensitive enzyme *R.EcoRI*.

TRDs but can also be functionally active when added to the ensemble of existing TRDs in both MTases M.ϕ3TII and M.(ϕ)BssHII. TRD0 can therefore be regarded as an independent modular element. Furthermore, these results highlight the great structural plasticity of both MTases to accommodate this TRD functionally at various positions. The addition of this TRD, however, interfered with the activity of the neighboring TRDs. Restriction analysis of plasmids encoding these chimeras revealed that the activities of TRD2 (*MluI*) in chimera V and TRDT (*TaqI*) in chimera IV were partially impaired.

Discussion

The post-replicative methylation of DNA at the C5 position of cytosines is found in a broad spectrum of organisms ranging from prokaryotes to human (Jost and Saluz, 1993). In prokaryotes, the major role of cytosine C5 methylation (like adenine A-N6 and cytosine C-N4 methylation) is to protect the genome against DNA-degrading nucleases (restriction/modification). In many eukaryotes, C5 methylation plays a pivotal role in development by regulating gene expression and possibly by safeguarding genome stability (Kass *et al.*, 1997; Yoder *et al.*, 1997).

The structure of MTases is remarkably conserved among various organisms. This is represented by a rigid succession of highly conserved regions and the presence of a variable region, determining enzyme specificity, at a defined location (Noyer-Weidner and Trautner, 1993; Bestor and Verdine, 1994). Evolution has affected primarily this variable region of the enzymes, generating MTases with a wide spectrum of specificities (Kumar *et al.*, 1994). Only a few variants of this canonical build up of MTases have been found. In one such variant type, M.*AquI*, the enzyme is divided into two separate subunits (Karreman *et al.*, 1990). In another, represented by the bacterial M.*BssHII* enzyme, domains show the conventional, but circularly permuted arrangement (Figure 5; Xu *et al.*, 1997). However, in all such enzymes, the spatial relationship between the various enzyme motifs remained the same, probably safeguarding the three-dimensional structure of prototype MTases with their division into a

Fig. 5. Circular representation of the build up of MTases. The circles show the conventionally structured M.*HhaI* enzyme, the permuted bacterial M.*BssHII*, the M.*AquI* enzyme which is formed by two separately expressed polypeptides, isoform I of M.(ϕ)BssHII and dnmt1 of the mouse, as an example of a eukaryotic MTase with an extensive N-terminal domain. The relative location of uninterrupted blocks of conserved elements I–VIII and IX–X and the TRD containing variable region 'V' are depicted in the inner circle. The two protein domains formed by the conserved (right) and variable (left) parts of the enzymes (according to Klimasauskas *et al.*, 1994; Reinisch *et al.*, 1995) are shaded. The N- and C-termini of the MTase polypeptides are marked by points and arrows. The drawing is not to scale.

catalytic domain and a target-recognizing domain (Malone *et al.*, 1995).

A totally different evolutionary event, generating a novel type of MTase, became evident with the identification of isoform I of *M.(ϕ)BssHII*. In this MTase, an extra TRD is placed at the N-terminus of the enzyme outside of the conventional ensemble of TRDs located between motifs VIII and IX (Figure 5). TRD0 was readily identified as a conventional TRD. Its amino acid sequence is compatible with those of the other TRDs of the MTase. In line with this, TRD0 is also active, if brought into the context of the variable region of *M.(ϕ)BssHII* or in the related enzyme *M. ϕ 3TII*.

The unconventional placement of TRD0 raises the question of how sequence-specific methylation is performed by this TRD. Whereas the interactions of TRDs of the variable region can be assumed to be similar to those described for TRDs of monospecific MTases, a different steric relationship between TRD, DNA and other enzyme regions involved in the methylation process must exist here. In view of the established three-dimensional structures of MTases folded into a large (catalytic) and small (recognition) domain, it is conceivable that the N-terminal TRD0 becomes placed at the join between these domains in the neighborhood of the variable region, and specific DNA recognition may occur in a similar way as with the TRDs of the variable region.

Since TRD0 can be transferred functionally into the variable region of the enzyme, the question arises as to why during evolution both TRD0 and the regulatory sequences needed for expression of the variant isoform I became located in an unusual position at the N-terminus of the protein. One possible explanation is that the mere enlargement of TRDs of the variable region would exceed a limit of TRDs tolerable in the variable region and would hence interfere with the proper function of the protein. Alternatively, additional functional properties of TRD0 itself, when inserted in the variable region, might impair the activity of neighboring TRDs, as observed in chimeras IV and V (Figures 1B and 3) in which the activity of the neighboring TRDs was reduced by the addition of TRD0. This interpretation is supported by the finding that in both chimeras the suppression of genuine TRD activities was less pronounced when the mutated, inactive form of TRD0 (P38S, see above) was inserted (Figures 3 and 4). Hence, the functional properties of TRD0 might have contributed to evolve the entirely new MTase structure in *M.(ϕ)BssHII* with TRD0 placed at an unusual position.

The identification and localization of TRD0 marks the N-terminus of MTases, in addition to the variable region, as a second area to which evolutionary flexibility can be assigned. Analyzing the known MTases, we find that only the ORFs of the related MTases *M.EcoRII*, *M.dcm* and *M.MlaX* extend significantly beyond the conserved domain I at their N-terminus. These parts of the ORFs have amino acid motifs which are also found in TRDs. These are, however, not likely to be used in target recognition. In the case of *M.EcoRII*, this sequence forms a DNA-binding domain involved in the autoregulation of the enzyme (Som and Friedman, 1994). The most prominent evolutionary event affecting the N-terminus is the generation of eukaryotic MTases by in-frame joining of the equivalent of a prokaryotic MTase ('catalytic domain') to a large

regulatory domain (Figure 5). It is apparent that this bulky regulatory domain does not affect the activity of the C-terminal catalytic domain. We do not know whether the prokaryotic MTases with extended N-termini represent evolutionary intermediates between prokaryotic and eukaryotic MTases. This possibility invites surveys of eukaryotic MTases for non-canonically placed TRDs in the N-terminal regulatory domain of such enzymes, especially in parts which have been described to discriminate between hemi- and unmethylated DNA.

Materials and methods

Reagents, bacteria and general techniques

Restriction endonucleases, T4 DNA ligase and *Taq* DNA polymerase were purchased from Boehringer Mannheim. All enzymes were used as recommended by the manufacturer. All molecular biological manipulations followed standard procedures (Sambrook *et al.*, 1989). All PCR-generated constructs were sequenced using the ABI PRISM™ *dRhodamine* Dye Terminator Cycle Sequencing Kit and the automated sequencing system (Applied Biosystems). Plasmids with the MTase genes were propagated in methylation-tolerant *E.coli* strains GM271 (obtained from M.G.Marinus; Krüger *et al.*, 1992) and NW22 [*F*+(*traD36-lacIq-lacZDM15-proAB*+) *hsdR2*(ori17), *mcrA*-, *mcrB*-1], which was from our strain collection.

MTase genes

Genes specifying the wild-type MTases *M.(ϕ)BssHII* together with its deletion derivative *M.(ϕ)BssHII Δ_{1-5}* and wild-type *M. ϕ 3TII*, respectively, have been described previously (Noyer-Weidner *et al.*, 1994; Schumann *et al.*, 1996). For the construction of chimeric *M.(ϕ)BssHII* MTases, the WT-V mutant (having an inactive TRD5, see Schumann *et al.*, 1996) was used. Chimeras I and II were constructed by fusing PCR products 5' and 3' of the respective ATGs of both MTase genes. In the course of the PCR-based construction of chimera II, an *Hae*II-sensitive mutant was identified with a point mutation at position C₂₅₉T leading to a P₃₈S amino acid change. Both the wild-type TRD0 and this mutated form (Mut38) of TRD0 were used for chimera constructions. Chimera IV and chimera IV_{Mut38} were constructed by inserting the appropriate PCR products (using primers #1 5'CGGAGATCTGAATTACGACAGG-ACTCCG3' and #2 5'ATCGGATCCCCGGGATGCCACGGCTGTCC3') amplified from the wild-type and the *M.(ϕ)BssHII* Mut38-encoding sequences, respectively, into the *Bgl*II site of the *M. ϕ 3TII* gene. The restriction sites [#1 (*Bgl*II), #2 (*Bam*HI)] contained in the primers used for cloning are underlined. To construct chimera III, chimera V and chimera V_{Mut38}, first the deletion derivatives of genes encoding *M.(ϕ)BssHII Δ_{1-5}* and the WT-V *M.(ϕ)BssHII* were generated by removing nucleotides A₁₅₄-G₂₆₂ (see Figure 1A) by a PCR-based mutagenesis. Subsequently, PCR-generated fragments corresponding to TRD0 and TRD0_{Mut38} (primers #1 and #2) were inserted into the *Bgl*II site of those deletion derivatives. Plasmids with genes encoding the His-tagged isoform I and II proteins were constructed by inserting appropriate PCR products into the *Nde*I and *Nsi*I sites of the derivative of a His tag overexpression vector pQE 60 (Qiagen). The fragment encoding His-tagged isoform I was generated from the wild-type (ϕ)*BssHII* sequence using primers #3 (5'GGCCATATGCGCGTCCGCGACCCCGCGG-AATTACGACAGG3') and #5 (5'CCAATGCATCTAGACCGCCTC-CTTCCG3'), while the isoform II-encoding fragment was generated with primers #4 (5'GGCCATATGCGCGTCCGCGACCCCGCGGTCG-ACGTAACG3') and #5. The primers #3 and #4 contain the restriction sites (underlined) followed by sequences encoding the IgA1 protease recognition site (italics) and the first five N-terminal amino acids (without initiator codons) from both isoforms, respectively. Since the presence of His tags in the fusion proteins did not affect their biological activity, the IgA1 protease treatment was not required to measure enzymatic activity.

MTase proteins

The His-tagged proteins were purified from cell extracts by one-step affinity chromatography on nickel-agarose columns (Qiagen) as suggested by the manufacturer. Western blotting of those proteins separated on 10% polyacrylamide gels was performed with polyclonal antibodies raised against unfractionated *M.(ϕ)BssHII* proteins. For N-terminal sequencing, the isoform I protein was purified from cells

carrying plasmid pBY-6 (Schumann *et al.*, 1996) which overexpresses both forms of M.(ϕ)BssHII. The cells were suspended in a buffer A (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA, 5% glycerol) and disrupted by a passage through the French press. The cell debris was removed by centrifugation in a Sorvall SS34 rotor at 18 000 r.p.m. for 30 min. The supernatant was applied to a heparin-Sepharose (LKB-Pharmacia) column equilibrated with buffer A. The proteins were eluted from the column by an NaCl gradient (100–600 mM) in buffer A. Fractions with the MTase proteins were pooled and dialyzed against buffer B (20 mM phosphate buffer, pH 6.8, 100 mM NaCl, 1 mM EDTA, 10 mM 2-mercaptoethanol, 5% glycerol) and loaded onto a phosphocellulose P11 (Whatman) column equilibrated with buffer B. The proteins were eluted using a linear (100–1000 mM) NaCl gradient in buffer B. Fractions containing the MTase proteins were subjected to a preparative (10% polyacrylamide) PAGE. The separated isoform I protein was blotted onto a PVDF membrane (Schleicher and Schuell) and subjected to Edman degradation analysis in an automated protein sequencer (Applied Biosystems). The amino acids determined were ANYDRTPQVPNQSRPVSAY.

DNA methylation assays

The *in vitro* methylation reactions were performed in a 5 mM Tris-HCl buffer (pH 7.5), 1 mM EDTA, 0.2 mM dithiothreitol and 0.64 mM SAM with 10 μ g of plasmid pMS119EH DNA and 0.5 μ g of purified protein. After 1 h incubation at 50°C in a final reaction volume of 100 μ l, the reactions were stopped by phenolization. The DNA was ethanol precipitated and dissolved in TE buffer. The susceptibility of the *in vitro* or *in vivo* methylated plasmid DNA to digestion by restriction enzymes was determined as described by Schumann *et al.* (1996). Digestion products were separated on 1% agarose gels. Resistance against digestion indicates modification of the appropriate restriction sites.

Acknowledgements

Discussions with Dr Xiaodong Cheng in the course of this work are greatly appreciated. We are grateful to R.Pawlik for performing the M.(ϕ)BssHII isoform I protein sequencing and to E.Philippi for her help with the artwork for this publication.

References

- Bestor, T.H. and Verdine, G.L. (1994) DNA methyltransferases. *Curr. Biol.*, **6**, 380–389.
- Jost, J.P. and Saluz, H.P. (1993) *DNA Methylation: Molecular Biology and Biological Significance*. Birkhäuser Verlag, Basel, Switzerland, Vol. 64.
- Karremans, C. and De, W.A. (1990) *Agmenellum quadruplicatum* M.AquI, a novel modification methylase. *J. Bacteriol.*, **172**, 266–272.
- Kass, S.U., Pruss, D. and Wolffe, A.P. (1997) How does DNA methylation repress transcription. *Trends Genet.*, **13**, 444–449.
- Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
- Kröger, T., Grund, C., Wild, C. and Noyer-Weidner, M. (1992) Characterization of the *mcrBC* region of *Escherichia coli* K12 wild-type and mutant strains. *Gene*, **114**, 1–12.
- Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J. and Wilson, G.G. (1994) The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.*, **22**, 1–10.
- Lauster, R., Trautner, T.A. and Noyer, W.M. (1989) Cytosine-specific type II DNA methyltransferases—a conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.*, **206**, 305–312.
- Leonhardt, H., Page, A.W., Weier, H.-U. and Bestor, T.H. (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, **71**, 865–874.
- Malone, T., Blumenthal, R.M. and Cheng, X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.*, **253**, 618–632.
- Mi, S. and Roberts, R.J. (1992) How M.MspI and M.HpaII decide which base to methylate. *Nucleic Acids Res.*, **20**, 4811–4816.
- Noyer-Weidner, M. and Trautner, T.A. (1993) Methylation of DNA in prokaryotes. In Jost, J.P. and Saluz, H.P. (eds), *Methylation of DNA in Prokaryotes*. Birkhäuser Verlag, Basel, Switzerland, Vol. 64, pp. 39–108.
- Noyer-Weidner, M., Walter, J., Terschüren, P.A., Chai, S. and Trautner, T.A. (1994) M. ϕ 3TII: a new monospecific DNA (cytosine-C5)

methyltransferase with pronounced amino acid sequence similarity to a family of adenine-N6-DNA-methyltransferases. *Nucleic Acids Res.*, **22**, 5517–5523.

- Reinisch, K.M., Chen, L., Verdine, G.L. and Lipscomb, W.N. (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell*, **82**, 143–153.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Schumann, J., Walter, J., Willert, J., Wild, C., Koch, D. and Trautner, T.A. (1996) M.BssHII, a multispecific cytosine-C5-DNA-methyltransferase with unusual target recognizing properties. *J. Mol. Biol.*, **257**, 949–959.
- Som, S. and Friedman, S. (1994) Regulation of EcoRII methyltransferase: effect of mutations on gene expression and *in vitro* binding to the promoter region. *Nucleic Acids Res.*, **22**, 5347–5353.
- Xu, S., Xiao, J., Posfai, J., Maunus, R. and Benner, J.N. (1997) Cloning of the BssHII restriction-modification system in *Escherichia coli*: BssHII methyltransferase contains circularly permuted cytosine-5 methyltransferase motifs. *Nucleic Acids Res.*, **25**, 3991–3994.
- Yoder, J.A. and Bestor, T.H. (1996) Genetic analysis of methylation patterns in plants and mammals. *Biol. Chem.*, **377**, 605–610.
- Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, **13**, 335–340.

Received March 25, 1999; revised and accepted April 29, 1999